JOINT NEAR-END LISTENING ENHANCEMENT AND FAR-END NOISE REDUCTION

Markus Niermann, Peter Jax, Peter Vary

Institute of Communication Systems
RWTH Aachen University, Germany
{niermann,jax,vary}@iks.rwth-aachen.de

ABSTRACT

Algorithms for Near-End Listening Enhancement (NELE) improve the intelligibility of speech from the far-end, played back in a near-end noisy environment, by adaptively filtering the speech signal and taking into account the near-end background noise characteristics. In contrast to previous works, this study considers that the speech from the far-end is also disturbed by additive noise. A noise reduction (NR) subsystem is concatenated to NELE and a joint control of NR and NELE is proposed. Listening tests confirm that the joint control leads to a significantly higher intelligibility and quality.

1. INTRODUCTION

In communication and announcement systems, speech signals from a communication partner at the far-end are often reproduced in a noisy near-end environment. Thus, the received signal may partly be masked by the near-end background noise. Near-End Listening Enhancement (NELE) is a technique to enhance the speech intelligibility and listening comfort [1, 2]. Since the near-end noise cannot be influenced by signal processing, NELE adaptively preprocesses the received far-end signal exploiting knowledge about the near-end noise. There are several methods for enhancement such as spectral weighting in subbands, adapted to the human auditory system [3, 4, 5, 6, 7, 8], and dynamic range compression [5]. In the literature, dynamical spectral weights are often determined by maximizing instantaneous intelligibility measures such as the Speech Intelligibility Index (SII) [3, 4, 8] or the mutual information [6, 7]. Several side constraints may be considered, e.g., upper bound for the total speech power, total amplification or power per sub-band. NELE was originally developed to be applied in mobile phones in the downlink. Further potential applications for NELE are hearing aids and public announcement systems, e.g., at railway stations. In the latter context, also the term "Speech Reinforcement" has been used [9].

Most previous works on NELE assume that the speech signal from the far-end is clean, i.e., unaffected by noise and other degradations. In many situations, however, this assumption is not appropriate. In the case of a telephone call between two mobile phones for example, the far-end speaker as well as the near-end listener might be located in noisy environments, thus the signal from the far-end is also degraded by noise. The authors of [7] propose a combination of beamforming for noise reduction (NR) and NELE by maximization of mutual information and conclude that a common consideration of near-end and far-end noise is necessary. In our contribution, we combine an SII-optimizing NELE system with a conventional, single-channel far-end NR system, analyze the interaction and come to the conclusion that the transmission of supplemental information between the NR and NELE systems is beneficial. Finally, we propose a solution for a combination of NR and NELE that considers the noise conditions at the near-end and far-end jointly. In the boundary cases without noise at the near-end or without noise at the far-end side, the algorithm behaves exactly as a conventional NR or as a conventional NELE system, respectively.

2. SYSTEM DESCRIPTION

Fig. 1 gives an overview of a communication system, consisting of a concatenation of far-end NR and near-end NELE subsystems. All processing takes place in the frequency domain. In mobile telephony, NR usually is applied at the transmitting device and NELE at the receiving device.

In this contribution, signals captured by microphones, for example $y(k)$ with sample index k, are segmented into frames with index λ, windowed by a $\sqrt{\mathrm{Hann}}$-Window and transformed to the frequency domain signal $Y_\mu(\lambda)$ with the frequency index μ by means of a Discrete Fourier Transform (DFT) of length M_F. The output signal $S_{\mathrm{enh},\mu}(\lambda)$ is transformed to the time domain by means of an inverse DFT, windowed by a $\sqrt{\mathrm{Hann}}$-Window and re-arranged via overlap-add prior to playing back at the near-end. In the figure, all signals are denoted in the frequency domain. For clarity, the transformations mentioned above as well as the frame index are omitted.

At the far-end, a microphone records a single-channel, noisy signal $Y_\mu = S_\mu + Q_\mu$. It consists of clean speech S_μ and additive noise Q_μ. The noisy far-end signal is processed by a NR system (Sec. 2.2) and a NELE system (Sec. 2.1). In the following, state-of-the-art systems for NR and for NELE are introduced and considered as a simple concatenation.

2.1. NELE

In this paper, a typical NELE algorithm that maximizes the Speech Intelligiblity Index (for simplicity without power constraint) is employed. Conventional NELE assumes that far-end noise is absent ($Q_\mu = 0$), and that NR is switched off ($\tilde{Y}_\mu = Y_\mu$). The far-end signal $y(k)$, which only contains speech in this case, is spectrally weighted with gains W_μ. The gains are determined based on a power estimate of the far-end signal $\tilde{Y}$ and the near-end disturbance D_μ. The disturbance is deduced from the near-end noise N_μ by incorporating spectral masking.

Internally, NELE works with bark-scale subbands, i.e. frequency-domain signals are transformed to the subband domain. The subbands are indexed as $i = 0, ..., M_{\mathrm{SB}} - 1$ and defined by a conversion matrix $\mathbf{C}$ with the elements $C_{i,\mu}$. Using the matrix, the averaged subband powers of speech and disturbance can be calculated by

$$\delta_{\tilde{\mathrm{Y}},i} = \sum_{\mu=0}^{M_F-1} C_{i,\mu} \mathop{\mathrm{E}}_{\mathrm{vad}}[|\tilde{Y}_\mu|^2], \qquad (1)$$

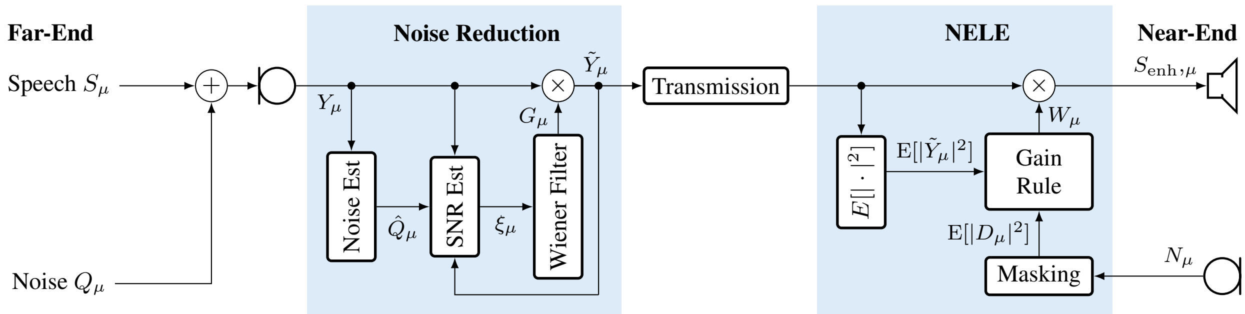

Fig. 1. Concatenation of noise reduction and a Near-End Listening Enhancement. All signals are sampled at 16 kHz and considered in the discrete frequency domain. A/D-converters, D/A-converters and frequency transforms are not shown.

$$\delta_{\mathrm{D},i} = \sum_{\mu=0}^{M_F-1} C_{i,\mu}\, \mathrm{E}[|D_\mu|^2], \tag{2}$$

where E is the linear short-time expectation operator which averages over time. In the implementation, the expectation operator is realized as autoregressive smoothing filter with the parameter β, e.g.,

$$\mathrm{E}[|A_\mu(\lambda)|^2] = \beta \cdot \mathrm{E}[|A_\mu(\lambda-1)|^2] + (1-\beta)\cdot|A_\mu(\lambda)|^2. \tag{3}$$

For speech estimation (1), a binary Voice Activity Detector (VAD) is used, since the estimate is updated only in time frames with speech activity. In the following, logarithmic powers are considered, e.g.,

$$L_{\tilde{\mathrm{Y}},i} = 10\cdot\log_{10}\delta_{\tilde{\mathrm{Y}},i}. \tag{4}$$

The disturbance level $L_{\mathrm{D},i}(\lambda)$ is limited, to prevent excessive amplifications in subbands with high noise levels.

$$L_{\mathrm{D}',i}(\lambda) = \min\left(L_{\mathrm{D},i}(\lambda), \mathrm{param}(\lambda)\right), \tag{5}$$

where $\mathrm{param}(\lambda)$ is determined such that the maximum level is at most $8\,\mathrm{dB}$ higher than the mean level [8].

We use a simplification of an SII optimizing NELE algorithm without power constraint [3]. The concept of the NELE rule is to ensure that the speech level in each subband is always sufficiently high to be intelligible, compared to the limited disturbance. The fixed parameter ΔL_{S} represents the necessary level difference. Moreover, an attenuation of the speech levels is not allowed. In total, the lowest possible weights $L_{\mathrm{W},i}$ are determined which satisfy both of the following conditions,

$$L_{\mathrm{W},i} + L_{\tilde{\mathrm{Y}},i} \geq L_{\mathrm{D}',i} + \Delta L_{\mathrm{S}}, \tag{6a}$$

$$L_{\mathrm{W},i} \geq 0\,\mathrm{dB}. \tag{6b}$$

The result is transferred to the amplitude domain, $W_i = 10^{L_{\mathrm{W},i}/20}$.

2.2. NELE with Far-End Noise Reduction

If background noise is present at the far-end, i.e. $Q_\mu \neq 0$, the following problem occurs: The NELE system interprets the noisy signal Y_μ as speech and will amplify not only the speech, but also the noise to overcome the masking threshold due to the near-end noise. The amplified noise disturbs the listener in addition to the near-end noise. Therefore, a state-of-the-art NR system is implemented. In a first approach, NR and NELE are simply concatenated without transmission of supplemental information between the two systems. Later, in Sec. 3, a joint solution is presented.

The NR block consists of noise estimation, signal-to-noise ratio (SNR) estimation and a Wiener Filter. A noise estimation algorithm, in this case the Speech Presence Probability algorithm [10], provides a noise estimate $\hat{Q}_\mu$ based on Y. Afterwards, the a-priori SNR ξ_μ is estimated using the Decision Directed Approach [11] and spectral gains are computed by means of a Wiener Filter,

$$G_\mu = \frac{\xi_\mu}{\xi_\mu + 1} \in [0,1]. \tag{7}$$

These gains are applied to the noisy far-end signal, $\tilde{Y}_\mu = G_\mu \cdot Y_\mu$, in order to maximize the SNR.

This simple solution succeeds in cancelling most of the far-end noise. However, problems occur:

- The Wiener Filter attenuates not only noise, but also speech. This leads to speech distortions if the far-end SNR is very low. This is nevertheless the preferred setting in typical NR scenarios without near-end noise. However, in the presence of near-end noise, sufficiently low levels of residual far-end noise can be partially masked by the near-end noise. Thus, it would be beneficial to set the NR less aggressive such that the far-end noise is not attenuated completely in this case. This reduces speech distortions significantly.
- In certain situations, NR and NELE influence the signal in opposite directions. This is the case when near-end listener and far-end speaker are both exposed to similar noise types. The NR strongly attenuates the noisy signal in the disturbed frequency range. Nevertheless, the NELE gain rule excessively amplifies the affected frequency range such that its level is ΔL_{S} above the near-end disturbance. Thus, the loudspeaker plays back amplified noise, distorted speech and musical tones.

3. PROPOSITION FOR JOINT APPROACH

For the derivation of a joint approach, the system of Fig. 2 is considered, which allows the exchange of information between the subsystems:

NELE → NR: Limit on NR aggressivity

NR → NELE: Levels of far-end speech and noise

3.1. Definition of Far-End SNRs

In this section, we derive equations to estimate the SNR of the far-end signal Y before NR and $\tilde{Y}$ after NR. Afterwards, we establish a rule to control either the output SNR or the noise attenuation. For the derivation and for a boundary experiment, it is assumed that speech S and noise Q are available separately. In real systems they have to be estimated on the basis of the noise estimation algorithm. The Wiener Filter gains G, however, are still calculated only on the basis of Y.

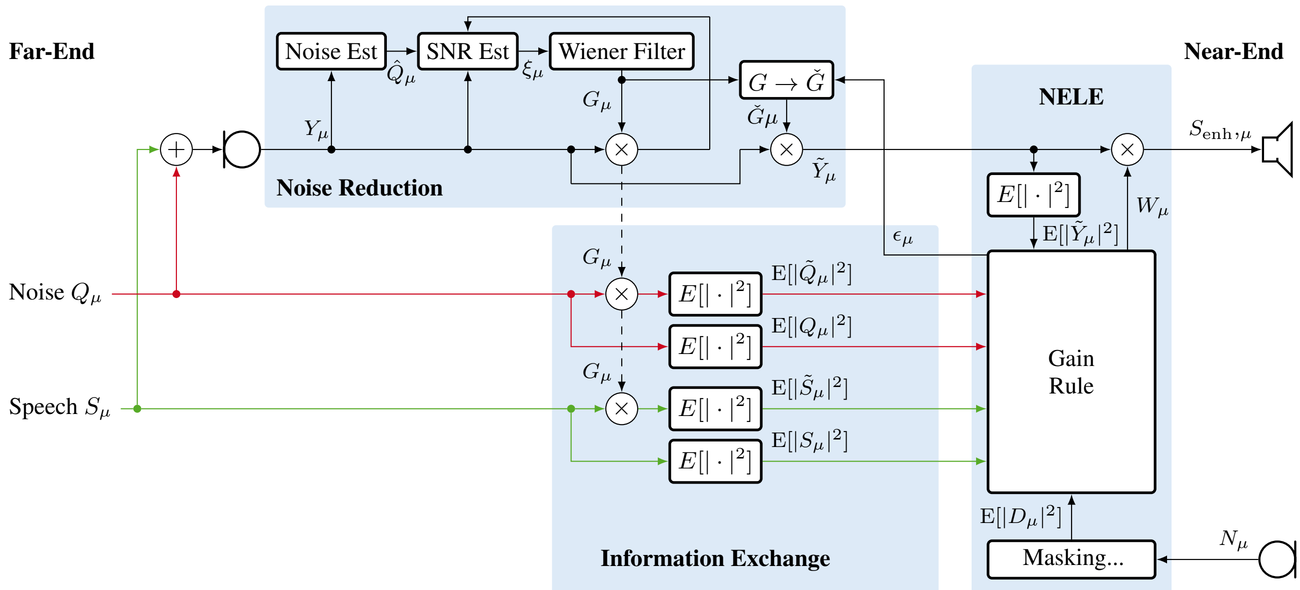

Fig. 2. Evaluation system for joint NR and NELE, allowing information exchange between both subsystems.

Due to the linearity of the gain multiplication, the processed far-end signal can be represented as the sum of the processed speech $\tilde{S}$ and processed noise $\tilde{Q}$:

$$G_\mu \cdot Y_\mu = G_\mu \cdot S_\mu + G_\mu \cdot Q_\mu = \tilde{S}_\mu + \tilde{Q}_\mu. \tag{8}$$

The phases of $\tilde{S}$ and $\tilde{Q}$ are denoted by $\varphi_{\tilde{S}}$ and $\varphi_{\tilde{Q}}$. A power estimation is performed using the short-term expectation operator E,

$$\begin{aligned}\mathrm{E}\left[|G_\mu \cdot Y_\mu|^2\right] =& \mathrm{E}\left[|\tilde{S}_\mu|^2\right] + \mathrm{E}\left[|\tilde{Q}_\mu|^2\right] \\ &+ 2 \cdot \mathrm{E}\left[|\tilde{S}_\mu| \cdot |\tilde{Q}_\mu| \cdot \cos\left(\varphi_{\tilde{S}}(\mu) - \varphi_{\tilde{Q}}(\mu)\right)\right].\end{aligned} \tag{9}$$

The last expectation term can be neglected (≈ 0) under the assumption that the processed speech and noise are uncorrelated.

The Wiener Filter in the DFT-domain is not able to increase the SNR locally, i.e. for a frequency bin. However, it can be increased globally, i.e. averaged over frequency subbands. Therefore, we specify the powers of speech $\delta^{\varepsilon}_{\mathrm{S},i}$ and noise $\delta^{\varepsilon}_{\mathrm{N},i}$ smoothed over time and averaged over all frequency bins μ that belong to the subband with index i. Initially, the superscript ε is a binary parameter that denotes if the power is estimated before NR ($\varepsilon = 0$) or after NR ($\varepsilon = 1$). The estimates are:

$$\delta^0_{\mathrm{S},i} = \sum_{\mu=0}^{M_F-1} C_{i,\mu} \cdot \mathrm{E}[|S_\mu|^2], \tag{10a}$$

$$\delta^0_{\mathrm{N},i} = \sum_{\mu=0}^{M_F-1} C_{i,\mu} \cdot \mathrm{E}[|Q_\mu|^2], \tag{10b}$$

$$\delta^1_{\mathrm{S},i} = \sum_{\mu=0}^{M_F-1} C_{i,\mu} \cdot \mathrm{E}[|\tilde{S}_\mu|^2] = \sum_{\mu=0}^{M_F-1} C_{i,\mu} \cdot \mathrm{E}[|S_\mu|^2 \cdot G^2_\mu], \tag{10c}$$

$$\delta^1_{\mathrm{N},i} = \sum_{\mu=0}^{M_F-1} C_{i,\mu} \cdot \mathrm{E}[|\tilde{Q}_\mu|^2] = \sum_{\mu=0}^{M_F-1} C_{i,\mu} \cdot \mathrm{E}[|Q_\mu|^2 \cdot G^2_\mu]. \tag{10d}$$

The far-end SNR with NR ($\varepsilon = 1$) and without NR ($\varepsilon = 0$) is

$$\mathrm{SNR}^{\varepsilon}_{\mathrm{f},i} = \frac{\delta^{\varepsilon}_{\mathrm{S},i}}{\delta^{\varepsilon}_{\mathrm{N},i}}, \tag{11}$$

where f denotes far-end. Usually, NR increases the SNR, i.e,

$$\mathrm{SNR}^0_{\mathrm{f},i} < \mathrm{SNR}^1_{\mathrm{f},i}. \tag{12}$$

Note that (12) may not be fulfilled if the noise estimate is imprecise.

3.1.1. *Control of Noise Reduction*

The Wiener Filter will, per definition, maximize the SNR. This may lead to undesirable speech distortions. Therefore, a method is deduced to reduce the aggressiveness and to achieve a specified SNR, which is between $\mathrm{SNR}^0_{\mathrm{f},i}$ and $\mathrm{SNR}^1_{\mathrm{f},i}$.

The control parameter ε_i, formerly a binary value, is now a soft value between zero and one that switches the NR softly on and off in order to control the aggressiveness. The values ε_i, $i = 0, ..., M_{\mathrm{SB}}-1$, in the subband-domain can be mapped to DFT-domain values ϵ_μ, $\mu = 0, ..., M_F - 1$. They are constant for all frequency bins μ that belong to one subband i.

The modified gain rule, which depends on ϵ_μ, is stated by

$$\breve{G}^2_\mu = 1 - \epsilon_\mu \cdot (1 - G^2_\mu), \quad \epsilon \in [0, 1]. \tag{13}$$

In the following, a relationship between ε_i and the target far-end $\mathrm{SNR}^{\varepsilon}_{\mathrm{f},i}$ is derived. Due to the linearity of the expectation operator, the speech power after spectral weighting with (13) can be written as

$$\delta^{\varepsilon}_{\mathrm{S},i} = \sum_{\mu=0}^{M_F-1} C_{i,\mu} \cdot \mathrm{E}[|S_\mu|^2 \cdot \breve{G}^2_\mu] \tag{14a}$$

$$\begin{aligned}&= (1-\varepsilon_i) \cdot \sum_{\mu=0}^{M_F-1} C_{i,\mu} \cdot \mathrm{E}[|S_\mu|^2] + \varepsilon_i \cdot \sum_{\mu=0}^{M_F-1} C_{i,\mu} \cdot \mathrm{E}[|S_\mu|^2 \cdot G^2_\mu] \\ &= (1-\varepsilon_i) \cdot \delta^0_{\mathrm{S},i} + \varepsilon \cdot \delta^1_{\mathrm{S},i}.\end{aligned} \tag{14b}$$

The same holds analogously for the noise,

$$\delta^{\varepsilon}_{\mathrm{N},i} = (1-\varepsilon_i) \cdot \delta^0_{\mathrm{N},i} + \varepsilon \cdot \delta^1_{\mathrm{N},i}. \tag{15}$$

This leads to an SNR after the modified spectral weighting of

$$\mathrm{SNR}_{\mathrm{f},i}^{\varepsilon} = \frac{\delta_{\mathrm{S},i}^{\varepsilon}}{\delta_{\mathrm{N},i}^{\varepsilon}} \tag{16a}$$

$$= \frac{(1-\varepsilon_i)\cdot\delta_{\mathrm{S},i}^{0} + \varepsilon_i\cdot\delta_{\mathrm{S},i}^{1}}{(1-\varepsilon_i)\cdot\delta_{\mathrm{N},i}^{0} + \varepsilon_i\cdot\delta_{\mathrm{N},i}^{1}}. \tag{16b}$$

Solving (16b) for ε, we obtain a rule to calculate ε for a target SNR,

$$\varepsilon_i = \frac{\mathrm{SNR}_{\mathrm{f},i}^{\varepsilon}\cdot\delta_{\mathrm{N},i}^{0} - \delta_{\mathrm{S},i}^{0}}{\mathrm{SNR}_{\mathrm{f},i}^{\varepsilon}\cdot\left[\delta_{\mathrm{N},i}^{0} - \delta_{\mathrm{N},i}^{1}\right] - \delta_{\mathrm{S},i}^{0} - \delta_{\mathrm{S},i}^{1}}, \tag{17}$$

which is clipped to take values between 0 and 1,

$$\tilde{\varepsilon}_i = \begin{cases} 0 & \text{if } \mathrm{SNR}_{\mathrm{f},i}^{\varepsilon} \le \mathrm{SNR}_{\mathrm{f},i}^{0} \\ 1 & \text{if } \mathrm{SNR}_{\mathrm{f},i}^{\varepsilon} \ge \mathrm{SNR}_{\mathrm{f},i}^{1} \\ \varepsilon_i & \text{else.} \end{cases} \tag{18}$$

Alternatively, we define the noise attenuation (NA) of the NR as

$$\mathrm{NA}_{\mathrm{f},i}^{\varepsilon} = \frac{\delta_{\mathrm{N},i}^{0}}{\delta_{\mathrm{N},i}^{\varepsilon}} \tag{19a}$$

$$= \frac{\delta_{\mathrm{N},i}^{0}}{(1-\varepsilon_i)\cdot\delta_{\mathrm{N},i}^{0} + \varepsilon_i\cdot\delta_{\mathrm{N},i}^{1}}. \tag{19b}$$

Solving (19b) for ε allows to calculate ε for a target NA, which has to be clipped to take values between 0 and 1,

$$\Rightarrow \tilde{\varepsilon}_i = \operatorname*{clip}_{0,1}\left(\frac{\mathrm{NA}_i - 1}{\mathrm{NA}_i}\cdot\frac{\delta_{\mathrm{N},i}^{0}}{\delta_{\mathrm{N},i}^{0} - \delta_{\mathrm{N},i}^{1}}\right). \tag{20}$$

In summary, we have derived methods to control the noise attenuation or the output SNR of the NR system.

3.2. Joint Control of NR and NELE

The modified NELE algorithm takes information from the NR into account. The new conditions for the joint control of the NELE parameter $L_{\mathrm{W},i}$ and the NR parameter ε_i with increasing priority are:

$$L_{\mathrm{W},i} + L_{\mathrm{S},i}^{\varepsilon} = L_{\mathrm{D}',i} + \Delta L_{\mathrm{S}}, \tag{21a}$$

$$L_{\mathrm{W},i} + L_{\mathrm{Q},i}^{\varepsilon} \le L_{\mathrm{D},i} + \Delta L_{\mathrm{N}}, \tag{21b}$$

$$L_{\mathrm{W},i} \ge 0\,\mathrm{dB}. \tag{21c}$$

Equation (21a) yields that the residual speech power after NR (with any value for ε) is not masked by the limited near-end disturbance. Condition (21b) restricts the residual far-end noise after NR to be masked partially or completely by the not-limited near-end disturbance $L_{\mathrm{D},i}$. Equation (21c) prohibits speech attenuations when the near-end noise is very low.

Solving (21a) for $L_{\mathrm{W},i}$ leads to weights of

$$L_{\mathrm{W},i} = L_{\mathrm{D}',i} + \Delta L_{\mathrm{S}} - L_{\mathrm{S},i}^{\varepsilon}. \tag{22}$$

The following steps are performed for each subband independently. First we check for $\varepsilon_i = 0$ whether condition (21c) is fulfilled and then we distinguish two cases:

- In the first case, if the condition is fulfilled, we have a set of two equations, (22) and (21b), and two unknown parameters, $L_{\mathrm{W},i}$ and ε. The weights from (22) are inserted into (21b), which leads to far-end SNRs that fulfill the conditions:

$$L_{\mathrm{S},i}^{\varepsilon} - L_{\mathrm{Q},i}^{\varepsilon} \ge \Delta L_{\mathrm{S}} - \Delta L_{\mathrm{N}} + \underbrace{L_{\mathrm{D}',i} - L_{\mathrm{D},i}}_{\le 0}. \tag{23}$$

  Since speech distortions should be as small as possible, we take the smallest SNR, that still fulfills (23), as target SNR,

$$\mathrm{SNR}_{\mathrm{f},i}^{\varepsilon} = 10^{\left(\Delta L_{\mathrm{S}} - \Delta L_{\mathrm{N}} + L_{\mathrm{D}',i} - L_{\mathrm{D},i}\right)/10}, \tag{24}$$

  and calculate the corresponding ε using (18). Knowing ε, $L_{\mathrm{S},i}^{\varepsilon}$ is calculated using (14b) and inserted into (22) to obtain the NELE weights. If the weights do not match with (21b), which means that the target SNR could not be reached, the weights are limited according to (21b).

- If (21c) is not fulfilled, we set $L_{\mathrm{W},i} = 0$ and deduce from (21b):

$$0 + L_{\mathrm{Q},i}^{\varepsilon} \le L_{\mathrm{D},i} + \Delta L_{\mathrm{N}} \tag{25}$$

$$\Rightarrow L_{\mathrm{Q},i}^{0} - L_{\mathrm{Q},i}^{\varepsilon} \ge L_{\mathrm{Q},i}^{0} - L_{\mathrm{D},i} - \Delta L_{\mathrm{N}}. \tag{26}$$

  The left-hand side represents the noise attenuation. In order to reduce speech distortions to a minimum, the lowest possible noise attenuation is chosen,

$$\mathrm{NA}_{\mathrm{f},i}^{\varepsilon} = 10^{\left(L_{\mathrm{Q},i}^{0} - L_{\mathrm{Q},i}^{\varepsilon} - \Delta L_{\mathrm{N}}\right)/10} = \frac{\delta_{\mathrm{Q},i}^{0}}{\delta_{\mathrm{D},i}^{0}\cdot 10^{\Delta L_{\mathrm{N}}/10}}, \tag{27}$$

  and ε is calculated for each subband by evaluating (20).

4. EVALUATION

The proprosed joint control (Sec. 3) has been evaluated using the evaluation system in Fig. 2 in comparison to the concatenation (Sec. 2.2) with the parameters as stated in Tab. 1. In an informal listening test, 16 participants listened to audio samples, simulated in three different noise scenarios, and were asked for their preference in terms of intelligibility and quality. The enhanced far-end signal has been rendered binaurally ($r = 2\,\mathrm{m}, \phi = 0°, \theta = 45°$) and mixed with binaurally recorded near-end noise. In 98 % of the cases, the quality of the joint control has been perceived to be better. With regard to intelligibility, 96 % of the answers indicate a preference for the joint control whereas 2 % indicate indifference. Measurements with the objective measure STOI [12] confirm this result.

Parameter	Settings
ΔL_{S}	10 dB
ΔL_{N}	0 dB
Sampling frequency f_{s}	16 kHz
FFT length M_F	512
Number subbands M_{SB}	21
Smoothing β	$\widehat{=}$ 2 seconds

Table 1. Algorithm parameters.

5. CONCLUSIONS

In this study, it has been shown that a joint control of far-end noise reduction and Near-End Listening Enhancement leads to significantly better results in terms of intelligibility and quality. This necessitates either a bi-directional information exchange between the far-end and near-end phone or setting up the NR at the receiver side instead of the transmitter side. If this is not possible, the near-end phone has to estimate the control parameters from the received signal, which is subject of further research.

6. REFERENCES

[1] W. B. Kleijn, J. B. Crespo, R. C. Hendriks, P. Petkov, B. Sauert, and P. Vary, "Optimizing Speech Intelligibility in a Noisy Environment: A unified view," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 43–54, 2015.

[2] M. Cooke, C. Mayo, C. Valentini-Botinhao, Y. Stylianou, B. Sauert, and Y. Tang, "Evaluating the intelligibility benefit of speech modifications in known noise conditions," *Speech Communication*, vol. 55, no. 4, pp. 572–585, May 2013.

[3] B. Sauert and P. Vary, "Near End Listening Enhancement Optimized with Respect to Speech Intelligibility Index," in *Proc. of European Signal Processing Conf. (EUSIPCO)*, New York, NY, Aug. 2009, pp. 1844–1848, Hindawi Publ.

[4] B. Sauert and P. Vary, "Near End Listening Enhancement Optimized with Respect to Speech Intelligibility Index and Audio Power Limitations," in *Proc. of European Signal Processing Conf. (EUSIPCO)*. Aug. 2010, pp. 1919–1923, EURASIP.

[5] T. Zorila, V. Kandia, and Y. Stylianou, "Speech-in-noise intelligibility improvement based on power recovery and dynamic range compression," in *Proc. of European Signal Processing Conf. (EUSIPCO)*, Aug. 2012, pp. 2075–2079.

[6] W. B. Kleijn and R. C. Hendriks, "A simple model of speech communication and its application to intelligibility enhancement," *IEEE Signal Processing Letters*, vol. 22, no. 3, pp. 303–307, 2015.

[7] S. Khademi, R. C. Hendriks, and W. B. Kleijn, "Jointly optimal near-end and far-end multi-microphone speech intelligibility enhancement based on mutual information," in *Proc. of IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*. 2016, pp. 654–658, IEEE.

[8] B. Sauert, *Near-End Listening Enhancement: Theory and Application*, PhD thesis, IND, RWTH Aachen University, Aachen, May 2014.

[9] J. B. Crespo and R. C. Hendriks, "Multizone speech reinforcement," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 22, no. 1, pp. 54–66, 2014.

[10] T. Gerkmann and R. C. Hendriks, "Noise power estimation based on the probability of speech presence," in *Proc. of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. 2011, pp. 145–148, IEEE.

[11] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 32, no. 6, pp. 1109–1121, Dec. 1984.

[12] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "A short-time objective intelligibility measure for time-frequency weighted noisy speech," in *Proc. of IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*. 2010, pp. 4214–4217, IEEE.